
DOCUMENT IMAGING

Service Description

Converting paper documents into electronic files helps us manage, store, access and archive the organizational information we have “locked up” in paper documents. This service utilizes high-quality document scanners, a state-of-the-art six engine Optical Character Recognition (OCR) system, and maintains quality controls to provide a successful Imaging solution. Once converted these electronic files can be indexed and searched, stored easier, and accessed and distributed faster, easier and cheaper than their paper originals.

Document Imaging service includes:

- Decollating documents to prepare them for feeding into sheet-fed scanners
- Scanning documents in black and white or color, 1 or 2-sided, up to 11”x17”
- Manual scanning of documents that cannot be fed into sheet-fed scanners
- Quality control to assure all pages are scanned and as readable as possible
- Running OCR software to create electronic files in various file formats: Acrobat Image + Text is the standard DOE document conversion format and Text, RTF and HTML formats are also available
- Delivering the final electronic files on CD’s
- Providing the original documents back in the same order as provided (though not re-stapled or re-bound)

To obtain specifications on preparing documents for processing and delivery to our production center please contact Helen Criares, (202) 586-5251, Ralph Freedman, (202) 586-2832, or see our web page at: <http://ma.mbe.doe.gov/admin/ImagingWorkflow.htm>

Pricing Policy

Pricing Method (Billing Calculation)

- Document Imaging charges are calculated per page based on usage.
 - Scanning and OCR - 28¢ per page.
 - These services are also offered a la carte, at:
 - Scanning Only - 20¢ per page.
 - OCR Only - 8¢ per page
 - All charges are calculated to be sufficient to cover the full business costs. Cost components included in this charge are:
 - Support service contract staff
 - Depreciation (equipment replacement costs)
 - Maintenance
 - Supplies

Billing Cycle

Document Imaging charges are billed monthly by usage. There are no charges for this service that are assessed annually or quarterly.

Availability of Detailed Usage Information

All job information is captured by the operator at the time of processing. This data is compiled into reports weekly and monthly. Information can be made available, on request, to include:

- Job requestor name
- Employee organization (division level)
- Date(s) of processing
- Number of pages processed
- Cost per job
- Files/documents processed

Business Line Objectives: Balanced Scorecard

- **Customers:** Provide customers with high-quality scanning and OCR in a timely and responsive manner.
- **Financials:** Ensure full cost recovery, and maintain competitive or better cost to the customer.
- **Internal Process:** Keep abreast of technology and processing enhancements to maintain best-of-class quality products.
- **Learning and Growth:** Keep the staff and the business line involved with the expanding world of electronic document technologies, services and products.

Service Standards

- Ensure that documents are maintained in as best condition as possible during the document preparation stage and throughout the processing. This includes maintaining the order of all pages in each document.
- Ensuring that all pages are scanned and OCR'd.
- Quality control to review documents to make sure they are as readable as possible.
- Work with customers to make sure the resulting electronic files follow naming conventions to facilitate storage and usage.
- Work with customers to identify documents that need to be manually scanned on flatbeds.
- Upon request, provide detailed usage reports.
- Provide further support on technical and policy issues relating to the resulting electronic files.

Points of Contact

Helen Criares, (202) 586-5251

Ralph Freedman, (202) 586-2832

For more information on document imaging services, please see our home page at <http://ma.mbe.doe.gov/admin/Imaging.htm>

Frequently Asked Questions About the Document Imaging Business Line

- Q.** What type of image quality can be expected from the scanning process?
- A.** High quality, highly readable images. Our scanners utilize Virtual Rescan™ technology to analyze each page and determine the best settings for brightness, contrast, and other

enhancements to ensure the best scan possible for both viewing the images as well as for the OCR processing. And our trained operators provide quality controls during the process to oversee the enhancements and make manual adjustments when needed.

Q. What file formats can the Optical Character Recognition (OCR) process create?

A. DOE's standard file format for document archival is Acrobat Image + Text. This is an Acrobat file that contains the scanned pages for viewing and printing purposes, retaining all information conveyed through the original printed page – charts, graphics, photos, handwriting on margins, etc. – with the OCR'd text behind the image so that the documents can be indexed and searched. This format is used by our Executive Secretariat and Freedom of Information and Privacy Act offices for official Departmental archiving, and is the standard format of the Federal Government for converting paper documents to electronic files for archival and access purposes.

The OCR process can convert scanned pages into other file formats including:

- ASCII (text-only)
- RTF for use in word processors. RTF format retains as much of the document formatting as possible but does not reverse-engineer paper documents into exact facsimiles of the originals.
- HTML for web-posting. HTML conversion does a good job of formatting the text, and can be useful in quickly posting documents on web sites, but it is still recommended that web programmers review and format the files before web posting.
- Acrobat Text Only. This is an Acrobat file that contains the OCR'd text but does not contain the images of the original scanned pages.
- Acrobat Image Only. This is a file that only contains the scanned pages, and can be opened with Acrobat.

For more detailed information about file formats, and indexing and searching files, please contact the Project Manager, Ralph Freedman, (202) 586-2832.

Q. What about the accuracy of the OCR processing, how reliable is it?

A. We utilize state-of-the-art OCR software that employs six OCR engines and artificial intelligence to vote on the accuracy of each character. This generates files that are significantly more accurate than any single OCR engine system could ever attain.

Q. Will the group clean up flagged character recognition from the OCR process?

A. This clean up work is not currently offered as part of this service. This is why we utilize a highly-accurate OCR system. If a customer needs 100% accuracy there are a few options available, and this would need to be discussed with the Project Manager, Ralph Freedman, (202) 586-2832.

Q. Other agencies and service bureaus provide this service. How does this Document Imaging service compare with them in quality and cost?

A. Extensive market research has been done in this area. The costs per page for this group is predominantly and significantly lower than any other organization that provides scanning and OCR. There are some service bureaus that can provide scanning only services at lower costs, which are mostly based on large volumes of documents, though they do not

offer the Virtual Rescan™ technology to adjust the images as part of their price. As far as getting six-engine OCR technology, service bureaus and GSA schedule providers tend to average double or more of our cost per page. Since DOE's service includes de-collating and otherwise preparing the documents for scanning, working with the customer on an as-needed basis for problem or unique documents, allowing for some manual scanning, allowing for some color scanning, can accept jobs that are both large and small, and does not require documents to leave the Headquarters complex, we have not found any service bureau or GSA schedule provider that can compete with our costs or quality. If you would like further information in this area, or would like to bring a service provider to our attention, please contact the Project Manager, Ralph Freedman, (202) 586-2832